



KARAKTERISTIK BUTIR SOAL TES LOGIKA MATEMATIKA DAN HIMPUNAN DITINJAU DARI TEORI TES KLASIK

¹⁾Syariful Fahmi*, ¹⁾Soffi Widyanesti Priwantoro, ¹⁾Anggit Prabowo

¹⁾ Pendidikan Matematika, Universitas Ahmad Dahlan Indonesia

syariful.fahmi@pmat.uad.ac.id

Abstract

Received :
13/02/2023

Accepted :
01/07/2025

Published :
13/08/2025

This study aims to analyze the formative test items for the Mathematical Logic and Sets course in the Digital Innovation Program of the Mathematics Education Study Program, FKIP UAD in terms of classical test theory. This research is descriptive research. The analysis was carried out on 20 multiple choice test items which were tested on 51 students. The analysis was carried out on the difficulty and discrimination index and reliability with the help of the ITEMAN (Item and Test Analysis) package program. The results of the analysis of the difficulty index showed that of the 20 items, there were 6 items that were difficult and 14 items that were moderate. In terms of discrimination, 18 items had good discrimination index and only 2 items had poor discrimination index. One of the causes of low discrimination index is caused by items that are too difficult. From these results, further revisions were made to the items that did not meet the good criteria. Reliability estimation using the Cronbach Alpha formula shows that the measurement results obtained through this test are reliable.

Keywords: items test, mathematics logic and set, classical test theory

Abstrak

Penelitian ini bertujuan untuk menganalisis karakteristik butir tes formatif mata kuliah Logika Matematika dan Himpunan dalam Program Inovasi Pembelajaran Digital Program Studi Pendidikan Matematika FKIP UAD ditinjau dari teori tes klasik. Penelitian ini merupakan penelitian deskriptif. Analisis dilakukan terhadap 20 butir tes pilihan ganda yang diujikan kepada 51 peserta tes. Analisis dilakukan terhadap indeks kesukaran, daya beda, dan reliabilitas dengan bantuan paket program ITEMAN (Item and Test Analysis). Hasil analisis ditinjau dari indeks kesukaran menunjukkan dari 20 butir soal, terdapat 6 butir yang sukar dan 14 butir yang sedang. Dari segi daya beda, 18 butir soal memiliki daya beda baik dan hanya 2 butir soal yang memiliki daya beda kurang baik. Salah satu penyebab kurang rendahnya daya beda disebabkan oleh butir soal yang terlalu sukar. Dari hasil ini selanjutnya dilakukan revisi terhadap butir-butir yang belum memenuhi kriteria baik. Estimasi reliabilitas menggunakan formula Cronbach Alpha menunjukkan bahwa hasil pengukuran yang diperoleh melalui tes ini reliabel.

Kata Kunci: butir soal, logika matematika dan himpunan, teori tes klasik

1. Pendahuluan

Sebagai upaya mendukung implementasi program “Merdeka Belajar-Kampus Merdeka” dan akselerasi pencapaian Indeks Kinerja Utama (IKU) Perguruan Tinggi, program studi Pendidikan Matematika FKIP UAD mengikuti Program Pengembangan dan Penyelenggaraan Inovasi Pembelajaran Digital (IPD) yang diinisiasi oleh Direktorat Pembelajaran dan Kemahasiswaan, Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan dan Kebudayaan. Kegiatan ini juga sebagai upaya untuk membekali mahasiswa keterampilan yang inovatif. Dalam kegiatan ini mendorong dosen dalam menghasilkan beragam model pembelajaran berbasis TIK yang dapat memfasilitasi

penyelenggaraan Merdeka Belajar, meningkatkan engagement, dan mampu menjawab tantangan era Industri 4.0 dan Society 5.0.

Program pengembangan dan penyelenggaraan Inovasi Pembelajaran Digital (IPD) yang diikuti program studi Pendidikan Matematika FKIP UAD pada periode ini terfokus pada mata kuliah Logika Matematika dan Himpunan. Mata kuliah ini terletak pada semester 1 dengan kredit 3 sks. Selama pandemic, pelaksanaan mata kuliah Logika Matematika dan Himpunan dilaksanakan secara daring berbasis Google Classrom, video pembelajaran melalui Youtube, serta social media whatsapp. Adapun materi pelajaran pada mata kuliah ini meliputi pernyataan atau proposisi, pernyataan majemuk dan tabel kebenarannya, tautologi, kontradiksi, pembuktian keabsahan argumen, himpunan, relasi, dan fungsi.

Untuk mengetahui keberhasilan pelaksanaan perkuliahan mata kuliah ini maka perlu dilakukan penilaian. Mengingat saat ini pelaksanaan perkuliahan masih sedang berjalan, maka penilaian yang dilakukan adalah penilaian formatif. Penilaian ini dilakukan untuk memberikan umpan balik kepada guru dan siswa selama pengajaran (Boston, 2002). Umpan balik yang diberikan sebagai bagian dari penilaian formatif membantu peserta didik menjadi sadar akan setiap kesenjangan yang ada antara tujuan yang diinginkan dan pengetahuan, pemahaman, atau keterampilan mereka saat ini dan membimbing mereka melalui tindakan yang diperlukan untuk mencapai tujuan (Ramaprasad, 1983; Sadler, 1989).

Penilaian formatif yang dilaksanakan pada mata kuliah Logika Matematika adalah menggunakan instrumen tes. Tes merupakan instrumen penilaian yang paling umum digunakan dalam dunia pendidikan (Adom, Mensah, & Dake, 2020). Selain sebagai instrumen, tes juga dapat dilihat sebagai prosedur standar yang digunakan untuk mengukur sampel perilaku secara sistematis dengan mengajukan serangkaian pertanyaan (Linn, 2008). Jenis tes yang dipilih adalah tipe pilihan ganda. Pemilihan format tes ini karena mampu mengukur tingkat berfikir yang kompleks (Haladyna, Rodriguez, & Stevens, 2019). Selain itu, tes pilihan ganda mampu mengukur konten dalam jumlah besar, lebih reliable, mudah diujicobakan, disimpan, digunakan, dan digunakan kembali, objektif, mudah diakomodasi berbagai teori tes, dan sebagian besar jenis konten dapat dites termasuk berfikir level tinggi (Haladyna, T. M. & Downing, 1989).

Butir tes pilihan ganda terdiri dari stem, yang mengidentifikasi pertanyaan atau masalah dan alternatif respons atau pilihan. Selanjutnya, siswa diminta memilih satu

alternatif yang terbaik melengkapi pernyataan atau menjawab pertanyaan (Clay, 2001). Butir-butir ini merupakan landasan dasar dari tes tersebut (Haladyna & Haladyna, 1992). Untuk itu, kualitas tes sangat penting untuk diperhatikan agar dapat menghasilkan hasil pengukuran yang akurat. Tes yang digunakan telah dikembangkan dengan dengan mengacu pada prosedur pengembangan tes mulai dari penyusunan tujuan tes, penyusunan kisi-kisi, menulis butir soal hingga validasi oleh ahli. Setelah dinyatakan valid, butir-butir soal selanjutnya dirakit menjadi sebuah perangkat tes yang dalam pelaksanaannya dilaksanakan secara online menggunakan platform Quizizz. Hasil tes berdasar data empirik selanjutnya perlu untuk dianalisis untuk melihat kualitas butir-butir soal ditinjau dari tingkat kesulitan, daya beda, dan reliabilitasnya. Dalam artikel ini akan dibahas kualitas butir-butir soal ditinjau dari pendekatan teori tes klasik.

2. Metode Penelitian

Penelitian ini merupakan penelitian deskriptif kualitatif. Penelitian dilakukan dengan menganalisis butir-butir soal tes formatif mata kuliah Logika Matematika dan Himpunan yang diselenggarakan sebagai bentuk implementasi Program Inovasi Pembelajaran Digital dari Direktorat Pembelajaran dan Kemahasiswaan, Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan dan Kebudayaan yang diselenggarakan oleh Program Studi Pendidikan Matematika FKIP UAD. Tes formatif yang diterapkan pada mata kuliah ini terdiri atas 20 butir soal pilihan ganda dengan 5 pilihan jawaban. Tes formatif dikerjakan oleh 51 mahasiswa. Tes dilaksanakan secara online dengan menggunakan platform online quiz, Quizizz. Hasil jawaban tes dianalisis dengan pendekatan teori tes klasik menggunakan bantuan paket program ITEMAN (Item and Test Analysis). Karakteristik tes yang dianalisis meliputi indeks kesukaran, daya beda, dan reliabilitasnya.

3. Hasil dan Pembahasan

Hasil penelitian menunjukkan hasil dari analisis indeks kesukaran, daya beda, dan reliabilitas instrument tes formatif Logika Matematika dan Himpunan Program Inovasi Pembelajaran Digital Program Studi Pendidikan Matematika FKIP UAD.

Tabel 1. Indeks Kesukaran Butir Soal

No. Soal	Indeks Kesukaran	Keterangan	No. Soal	Indeks Kesukaran	Keterangan
1.	0,70	Sedang	11.	0,275	Sukar
2.	0,588	Sedang	12.	0,294	Sukar

No. Soal	Indeks Kesukaran	Keterangan	No. Soal	Indeks Kesukaran	Keterangan
3.	0,137	Sukar	13.	0,470	Sedang
4.	0,627	Sedang	14.	0,345	Sedang
5.	0,70	Sedang	15.	0,433	Sedang
6.	0,608	Sedang	16.	0,236	Sukar
7.	0,216	Sukar	17.	0,490	Sedang
8.	0,373	Sedang	18.	0,157	Sukar
9.	0,510	Sedang	19.	0,470	Sedang
10.	0,529	Sedang	20.	0,336	Sedang

Hasil analisis indeks kesukaran selanjutnya dikategorikan berdasar kriteria bahwa jika indeks kesukarannya kurang dari 0,3 maka terkategori sukar, 0,3 sampai dengan 0,7 terkategori sedang, dan lebih dari 0,7 terkategori mudah. Berdasarkan Tabel 1 dapat dilihat bahwa dari 20 butir soal, terdapat 6 butir yang terkategori sukar (butir nomor 3, 7, 11, 12, 16, dan 18) dan 14 butir lainnya terkategori sedang. memiliki tingkat kesukaran “sedang”. Butir-butir tersebut baik mengingat indeks kesukaran yang diharapkan adalah berkisar antara 0,3 sampai dengan 0,7 (Mardapi, 2018) Sehingga berdasarkan kriteria tingkat kesukaran maka soal dapat dikatakan baik.

Tabel 2. Daya Beda Butir Soal Pilihan Ganda

No. Butir	Daya beda	Keterangan	No. Butir	Daya beda	Keterangan
1.	0,596	Baik	11.	0,507	Baik
2.	0,713	Baik	12.	0,337	Baik
3.	0,604	Baik	13.	0,470	Baik
4.	0,488	Baik	14.	0,345	Baik
5.	0,130	Tidak baik	15.	0,433	Baik
6.	0,417	Baik	16.	0,236	Baik
7.	0,326	Baik	17.	0,490	Baik
8.	0,243	Baik	18.	0,157	Tidak baik
9.	0,627	Baik	19.	0,470	Baik
10.	0,674	Baik	20.	0,336	Baik

Menurut Fernandes (1984) daya yang baik adalah di atas 0,2. Berdasar Tabel 2, dari 20 butir soal hanya ada dua butir soal yang daya bedanya tidak baik (butir nomor 5 dan 15). Butir-butir lainnya memiliki daya beda yang baik. Untuk estimasi reliabilitas, dilakukan dengan menggunakan formula Cronbach Alpha. Hasil estimasi menunjukkan koefisien reliabilitas tes sebesar 0,6.

Program Inovasi Pembelajaran Digital (IPD) merupakan program yang diselenggarakan oleh Direktorat Pembelajaran dan Kemahasiswaan Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi Kementerian Pendidikan, Kebudayaan, Riset,

dan Teknologi. Program ini diikuti oleh Program Studi Pendidikan Matematika FKIP UAD bekerja sama dengan Universitas PGRI Yogyakarta, dan Universitas Muhammadiyah Jember. Program IPD dilaksanakan pada mata kuliah Logika Matematika dan Himpunan.

Saat ini perkuliahan memasuki minggu ketujuh. Sebagai bentuk penilaian dalam perkuliahan ini maka diadakan penilaian formatif dengan menggunakan instrumen tes. Instrumen tersebut dalam pengembangannya mengacu pada 12 langkah prosedur pengembangan instrumen tes yang dikembangkan oleh Downing & Haladyna (2006) yang berturut-turut meliputi perencanaan, pendefinisian konten, menentukan spesifikasi tes, pengembangan butir, merakit tes, memproduksi tes, pengadministrasian tes, pensekoran, menentukan passing score, melaporkan hasil tes, item banking, dan laporan teknis tes.

Pada langkah keenam dan ketujuh pengembangan tes ini, dilakukan validasi ahli untuk melihat kesesuaian konten yang diukur. Validasi dilakukan oleh rekan sejawat yang ahli sesuai bidang mata kuliah dan yang ahli di bidang evaluasi pembelajaran. Hasil analisis oleh ahli menunjukkan bahwa instrumen yang dikembangkan sudah memenuhi kriteria validitas ditinjau dari kesesuaian konten (content validity). Selanjutnya, pada langkah kedelapan, Scoring test responses dilakukan analisis butir soal. Analisis yang dilakukan didasarkan pada teori tes klasik untuk melihat indeks kesukaran, daya beda, dan reliabilitas tes. Analisis karakteristik-karakteristik ini penting dalam pengembangan tes (Haladyna, 2004).

Ditinjau dari indeks kesukaran butir soal, soal yang baik adalah soal yang memiliki indeks kesukaran sedang. Indeks kesukaran merupakan ukuran proporsi total peserta ujian yang menjawab item dengan benar dan paling sering disebut sebagai nilai-p (Musa, Shaheen, & Elmardi, 2018). Indeks kesukaran berkisar antara 0 sampai dengan 1. Adapun indeks kesukaran yang ideal adalah antara 0,3 sampai dengan 0,7. Jika kurang dari 0,3 maka butir soal tersebut terlalu sulit dan di atas 0,7 maka butir tersebut terlalu mudah. Butir yang baik adalah jika butir soal tersebut tidak terlalu mudah dan tidak terlalu sulit (Tavakol & Dennick, 2011b). Hasil analisis butir menunjukkan 14 butir terkategori baik karena dalam kriteria sedang. Dari 6 butir yang sulit, butir yang paling sulit adalah butir nomor 3.

3. Negasi dari : Jika langit mendung maka hari hujan
- ☐ a) Jika langit tidak mendung maka hari tidak hujan ☐ b) Jika langit mendung maka hari tidak hujan
- ☐ c) Langit tidak mendung dan hari tidak hujan ☐ d) Langit mendung atau hari hujan
- ☐ e) Langit mendung dan hari tidak hujan

Gambar 1. Butir soal paling Sulit

Berdasar hasil analisis pada Gambar 1, soal ini terlalu sulit karena dalam prosedur penyelesaiannya peserta tes harus merubah ke pernyataan majemuk terlebih dahulu baru kemudian menegasikan pernyataan majemuk tersebut. Ilustrasinya adalah sebagai berikut.

$$p \rightarrow q \equiv \sim p \vee q \equiv p \wedge \sim q$$

Dengan demikian, jawabannya adalah e) Langit mendung dan hari tidak hujan. Karena soal itu terlalu sulit, maka sebagian besar siswa tidak dapat mengerjakan soal tersebut dengan benar. Pada akhirnya tidak mampu membedakan mana peserta tes yang berkemampuan tinggi dan rendah. Revisi yang dilakukan terhadap butir ini adalah dengan merubah stem menjadi: Negasi langit mendung atau hari hujan adalah Jawabannya adalah langit tidak mendung dan hari tidak hujan. Revisi juga dilakukan terhadap butir-butir soal nomor 7, 11, 12, 16, dan 18.

Ditinjau dari daya bedanya, terdapat dua butir yang daya bedanya kurang baik, butir nomor 5 dan 18. Butir nomor 5 terindikasi memiliki daya beda rendah karena butir ini cenderung terlalu mudah bagi peserta tes. Hal yang sebaliknya terjadi pada butir nomor 18. Butir ini terlalu sukar sehingga tidak mampu membedakan peserta tes bekemampuan tinggi dengan peserta tes dengan kemampuan rendah. Dengan demikian, revisi yang dilakukan adalah dengan merubah soal sehingga tingkat kesulitannya menjadi sedang sehingga mampu membedakan peserta tes. Dengan mendesain tes yang tingkat kesulitannya sedang akan meningkatkan daya beda butir soal (Sim & Rasiah, 2006). Untuk meningkatkan reliabilitas secara otomatis merupakan dampak dari revisi butir-butir soal yang telah dilakukan. Hasil studi Tavakol & Dennick (2011a), revisi yang dilakukan terhadap butir-butir yang belum sesuai akan berdampak pada reliabilitas tes.

4. Kesimpulan dan Saran

Hasil penelitian menunjukkan hasil analisis karakteristik butir tes formatif mata kuliah Logika Matematika dan Himpunan sebanyak 20 butir soal pilihan ganda dengan 5 pilihan. Ditinjau dari indeks kesukaran menunjukkan dari 20 butir soal, terdapat 6 butir

yang sukar dan 14 butir yang sedang. Dari segi daya beda, 18 butir soal memiliki daya beda baik dan hanya 2 butir soal yang memiliki daya beda kurang baik. Penyebab kurang rendahnya daya beda disebabkan oleh butir soal yang terlalu sukar dan terlalu mudah. Dari hasil ini selanjutnya dilakukan revisi terhadap butir-butir yang belum memenuhi kriteria baik. Untuk estimasi reliabilitas menunjukkan bahwa hasil pengukuran tes formatif menunjukkan hasil yang reliabel dan diperkirakan akan meningkat seiring revisi yang telah dilakukan pada butir-butir yang terlalu sukar dan terlalu mudah.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Direktorat Belmawa Dirjen Dikti Kemdikbud yang telah mendanai kegiatan Hibah Inovasi Pembelajaran Digital. Ucapan terima kasih juga disampaikan Universitas Ahmad Dahlan yang telah memberikan dukungan atas terlaksananya Inovasi Pembelajaran Digital yang diselenggarakan oleh Program Studi Pendidikan Matematika FKIP UAD.

Pustaka

- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education*, 9(1), 109–119. <https://doi.org/10.11591/ijere.v9i1.20457>
- Boston, C. (2002). The Concept of Formative Assessment. *Practical Assessment, Research, and Evaluation*, 8(2002), 9.
- Clay, B. (2001). *Is This a Trick Question ? Is This a Trick*. Kansas: Kansas Curriculum Center.
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of Test Development*. Mahwah: Lawrence Erlbaum Associates.
- Fernandes, H. J. X. (1984). *Testing and Measurement*. Jakarta: Planning, Evaluation, and Development.
- Haladyna, T. M. & Downing, S. . (1989). A Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2(1), 37–50.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haladyna, T. M., & Haladyna, T. M. (1992). The Effectiveness of Several Multiple-Choice Formats, 5(1), 73–88. <https://doi.org/10.1207/s15324818ame0501>
- Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are Multiple-choice Items Too Fat ? *Applied Measurement in Education*, 32(4), 350–364. <https://doi.org/10.1080/08957347.2019.1660348>
- Linn, R. L. (2008). *Measurement and assessment in teaching*. India: Pearson Education.

-
- Mardapi, D. (2018). Teknik penyusunan instrumen tes dan nontes. Yogyakarta: Mitra Cendekia Offset.
- Musa, A., Shaheen, S., & Elmardi, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University Abstract : Khartoum Medical Journal, 11(2), 1477–1468.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Sim, S. M., & Rasiah, R. I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals of the Academy of Medicine Singapore*, 35(2), 67–71.
- Tavakol, M., & Dennick, R. (2011a). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tavakol, M., & Dennick, R. (2011b). Post-examination analysis of objective tests. *Medical Teacher*, 33(6), 447–458. <https://doi.org/10.3109/0142159X.2011.564682>